OPIC
OFFICE DE LA PROPRIÉTÉ INTELLECTUELLE DU CANADA

CIPO
CANADIAN INTELLECTUAL PROPERTY OFFICE

BH

Ottawa Hull K1A 0C9

(21) (A1) 2,108,113
(22) 1993/10/08
(43) 1995/04/09

(51) INTL.CL.[5] C12N-015/52; C12N-015/76; C12N-009/00; C12P-017/18; C12N-001/21

(19)(CA) **APPLICATION FOR CANADIAN PATENT** (12)

(54) DNA Sequence Encoding Enzymes of Clavulanic Acid Biosynthesis

(72) Jensen, Susan E. - Canada ;
Aidoo, Kwamena A. - Canada ;
Paradkar, Ashish S. - Canada ;

(71) Governors of the University of Alberta (The) - Canada ;

(57) 39 Claims

DNA SEQUENCE ENCODING ENZYMES OF CLAVULANIC ACID BIOSYNTHESIS

This invention relates to methods for the production of the antibiotic, clavulanic acid.

Background of the Invention

Clavulanic acid is a broad spectrum beta-lactamase inhibitor and is an important antibiotic for the treatment of infectious diseases. It is produced commercially by the gram-positive mycelial prokaryote Streptomyces clavuligerus, which also produces the β-lactam antibiotics penicillin N, desacetoxy cephalosphorin C and cephamycin C. Until recently, however, the pathway employed for clavulanic acid biosynthesis was much less well understood than the pathways leading to these other antibiotics.

Without knowledge of the pathway for clavulanic acid biosynthesis, it was not possible to isolate the genes coding for the key enzymes and to manipulate these genes to increase antibiotic yield or permit production of the antibiotic in heterologous systems.

One of the earliest enzymes of the pathway to be purified and characterised was clavaminic acid synthase. Two isozymes have now been identified and characterised (Marsh et al., (1992), Biochem., vol. 31, pp. 12648-657).

European Patent Application 0349121 describes a DNA restriction fragment encoding a portion of the genetic information involved in clavulanic acid synthesis but provides no sequence information.

Until the work of the present inventors, the complete complement of genes required for clavulanic acid synthesis had not been identified. The present inventors have now isolated, cloned and sequenced an 11.6 kb genomic DNA sequence from S. clavuligerus which codes for eight proteins and enables the production of clavulanic

Figure 7 shows an alignment of the amino acid sequence of CLA (S. clavuligerus CLA) with those of E. Coli agmatine ureohydrolase (E. Coli AUH), yeast arginase (yeast ARG), rat arginase (rat ARG) and human arginase (human ARG).

Figure 8 shows a Southern blot of NcoI digests of genomic DNA from five presumptive mutants (lanes 1-5) and from wild-type S. clavuligerus (lane 6). Panel A : membranes probed with cla-specific probe. Panel B : membranes probed with tsr-specific probe.

Figure 9 shows restriction enzyme maps of S. clavuligerus DNA inserts in cosmids. A. Restriction enzyme map of cosmid K6L2. B. Partial restriction enzyme map of cosmid K8L2. C. Restriction map of cosmids K6L2 and K8L2 indicating location of pcbC gene in relation to cla. D. The 2.0 kb NcoI fragment encompassing the cla gene used in generating nested deletions for sequencing. Abbreviations: Ba, BamHI; B,BglII; E,EcoR1; K,KpnI; N, NcoI; S,SalI; and Sm,SmaI.

Figure 10 shows the deduced amino acid sequence (Sequence ID No.:3) of ORF1 of Figure 2.

Figure 11 shows the deduced amino acid sequence (Sequence ID No.:4) of ORF2 of Figure 2.

Figure 12 shows the deduced amino acid sequence (Sequence ID No.:5) of ORF3 of Figure 2.

Figure 13 shows the deduced amino acid sequence (Sequence ID No.:6) of ORF4 of Figure 2.

Figure 14 shows the deduced amino acid sequence (Sequence ID No.:7) of ORF5 of Figure 2.

Figure 15 shows the deduced amino acid sequence (Sequence ID No.:8) of ORF6 of Figure 2.

Figure 16 shows the deduced amino acid sequence (Sequence ID No.:9) of ORF7 of Figure 2.

Figure 17 shows the deduced amino acid sequence (Sequence ID No.:10) of ORF8 of Figure 2.

Figure 18 shows the deduced amino acid sequence (Sequence ID No.:11) of ORF9 of Figure 2.

when introduced into the non-clavulanate producer S. lividans as described in Example 4, enabled that species to produce clavulanic acid. This indicates that the 11.6 kb fragment contains all the genetic information required for clavulanate production.

As will be understood by those skilled in the art, the identification of the DNA sequence encoding the enzymes required for clavulanate synthesis will permit genetic manipulations to modify or enhance clavulanate production. For example, clavulanate production by S. clavuligerus may be modified by introduction of extra copies of the gene or genes for rate limiting enzymes or by alteration of the regulatory components controlling expression of the genes for the clavulanate pathway.

Heterologous organisms which do not normally produce clavulanate may also be enabled to produce clavulanate by introduction, for example, of the 11.6 kb DNA sequence of the invention by techniques which are well known in the art, as exemplified herein by the production of S. lividans strains capable of clavulanate synthesis. Such heterologous production of clavulanic acid provides a means of producing clavulanic acid free of other contaminating clavams which are produced by S. clavuligerus.

Suitable vectors and hosts will be known to those skilled in the art; suitable vectors include pIJ702, pJOE829 and pIJ922 and suitable hosts include S. lividans, S. parvulus, S. griseofulvus, S. antibioticus and S. lipmanii.

Additionally, the DNA sequences of the invention enable the production of one or more of the enzymes of the clavulanate pathway by expression of the relevant gene or genes in a heterologous expression system.

The DNA sequences coding for one or more of the pathway enzymes may be introduced into suitable vectors and hosts by conventional techniques known to those skilled in the art. Suitable vectors include pUC118/119

shown in Figure 3. ORF 4 corresponds to cla. ORF 1,7 & 8 are oriented in the opposite direction to pcbC. ORFs 2-6 and ORF 10 are all oriented in the same direction as pcbC. ORFs 2 and 3, and ORFs 4 and 5 are separated by very short intergenic regions suggesting the possibility of transcriptional and translational coupling. Table 1 summarises the nucleotide sequences and lengths of ORFs 1-10.

When the predicted amino acid sequences of proteins encoded by ORFs 1 - 10 were compared to protein sequence databases, some similarities were noted in addition to the already mentioned similarity between CLA and enzymes of arginine metabolism. ORF 1 showed a low level of similarity to penicillin binding proteins from several different microorganisms which are notable for their resistance to β-lactam compounds.

An EcoRI fragment of the 15 kb DNA sequence, containing 11.6 kb DNA, was cloned into a high copy number shuttle vector and introduced into S. lividans, as described in Example 4. Of seventeen transformants examined, two were able to produce clavulanic acid, indicating that the 11.6 kb fragment contains all the necessary genetic information for clavulanic acid production.

This 11.6 kb fragment encompasses ORF 2 to ORF 9 of the 15 kb DNA sequence.

ORF 2 shows a high degree of similarity to acetohydroxyacid synthase (AHAS) enzymes from various sources. AHAS catalyses an essential step in the biosynthesis of branched chain amino acids. Since valine is a precursor of penicillin and cephamycin antibiotics, and valine production is often subject to feedback regulation, it is possible that a deregulated form of AHAS is produced to provide valine during the antibiotic production phase. Alternatively, an AHAS-like activity may be involved in clavulanic acid production. While the presently recognized intermediates in the clavulanic acid

EXAMPLES

Example 1

Bacterial strains, vectors and growth conditions.

Streptomyces clavuligerus NRRL 3585, Streptomyces jumonjinenisis NRRL 5741, Streptomyces lipmanii NRRL 3584, Streptomyces griseus NRRL 3851, Nocardia lactamdurans NRRL 3802 and Streptomyces cattleya NRRL 3841 were provided by the Northern Regional Research Laboratories, Peoria, Il. Streptomyces antibioticus ATCC 8663 and Streptomyces fradiae ATCC 19609 were obtained from the American Type Culture Collection, Rockville, MD. Streptomyces lividans strains 1326 and TK24 were provided by D.A. Hopwood (John Innes Institute, Norwich, U.K.), Streptomyces venezuelae 13s and Streptomyces griseofuscus NRRL B-5429 were obtained from L.C. Vining (Department of Biology, Dalhousie University, Halifax, N.S.). Cultures were maintained on either MYM (Stuttard (1982) J. Gen. Microbiol., v. 128, pp. 115-121) or on a modified R5 medium (Hopwood et al. (1985) in "Genetic Manipulation of Streptomyces : a laboratory manual", John Innes Foundation, U.K.) containing maltose instead of glucose and lacking sucrose (R5-S). Escherichia coli MV1193 (Zoller and Smith (1987) Methods in Enzymology, v. 154, pp. 329-349), used as recipient for all of the cloning and subcloning experiments, was grown in Luria Broth (LB; Sambrook et al. (1989) in "Molecular Cloning : a laboratory manual", Cold Spring Harbour, N.Y.) or on LB agar (1.5%) plates containing ampicillin (50 μg/mL) or tetracycline (10 μg/mL). The cloning vectors pUC118 and pUC119 (Vieira and Messing (1987) Methods in Enzymology, v. 153, pp. 3-11) were provided by J. Vieira (Waksman Institute of Microbiology, Rutgers University, Piscataway, N,J.). The plasmid vector pJOE829 was generously provided by J. Altenbuchner (University of Stuttgart, Stuttgart, Germany). The plasmid pIJ702 was obtained from the American Type Culture Collection,

The probe was designed as an 8-fold degenerate mixture of oligonucleotides to take into consideration the biased codon usage of Streptomyces (Bibb et al., 1984, Wright and Bibb (1992), Gene, v. 113, pp. 55-65).). End-labelled probe was then used to screen a cosmid library of S. clavuligerus genomic DNA fragments as described in Materials and Methods.

A library of S. clavuligerus genomic DNA fragments (15-22 kb size fractionated fragments) was constructed as previously described (Doran et al. (1990), J. Bacteriol., v. 172, pp. 4909-4918). using the cosmid vector pLAFR3. A collection of 1084 isolated E. coli colonies containing recombinant cosmids was screened for the presence of cla using the 24-mer mixed oligonucleotide probe (Fig. 1) which had been end-labelled with [γ-^{32}P]dATP and polynucleotide kinase (Boehringer Mannheim). Colony hybridization and subsequent washing was performed as described by Sambrook et al., (1989), at 55°C with a final wash in 0.2X SSC (IX SSC, 0.15M NaCl and 0.015M sodium citrate) and 0.1% SDS.

Five colonies which gave strong hybridization signals were isolated from the panel of 1084 clones, and restriction analysis showed that the positive clones contained overlapping fragments of DNA. Two clones, K6L2 and K8L2, with sequences that spanned about 40 kb of the S. clavuligerus genome, were chosen for further analysis. Clone K8L2 contained about 22 kb of S. clavuligerus genomic DNA and included a portion of cla and all of the pcbC gene which encodes IPNS in the penicillin/cephamycin biosynthetic pathway. A restriction map of K6L2 is shown in Fig. 9. Within the approximately 27 kb of DNA contained in K6L2, the oligonucleotide probe hybridized to a 2.0 kb NcoI fragment which was subsequently found to contain the entire cla gene. Hybridization studies, restriction mapping and DNA sequence analysis revealed that cla was situated 5.67 kb downstream of the pcbC gene of S. clavuligerus (Fig. 9).

program described above. The AUH sequence had previously been aligned with the three ARG sequences (Szumanski & Boyle (1990), J. Bacteriol., v. 172, pp. 538-547). Identical matches in two or more sequences are indicated with upper case letters.

Example 2

DNA hybridization

Genomic DNA preparations from various Streptomyces species were isolated as described by Hopwood et al. (1985). For interspecies DNA hybridization analysis, 2.0 μg amounts of genomic DNA preparations were digested with NcoI for 16h, and electrophoresed in 1.0% agarose gels. The separated DNA fragments were then transferred onto nylon membranes (Hybond-N, Amersham) and hybridized with a cla specific probe prepared by labeling an internal 459 bp SalI fragment (Fig. 1) with [α-^{32}P]dATP by nick translation. Hybridization was done as described by Sambrook et al., (1989). Hybridization membranes were washed twice for 30 min in 2X SSC; 0.1% SDS and once for 30 min in 0.1X SSC; 0.1% SDS at 65°C.

Sequences homologous to cla in other Streptomycetes

Three of six producers of β-lactam antibiotics, S. clavuligerus, S. lipmanii and S. jumonjinensis showed positive hybridization signals whereas S. cattleya, S. griseus, and N. lactamdurans did not (data not shown). None of the nonproducing strains examined, S. venezuelae, S. lividans, S. fradiae, S. antibioticus and S. griseofuscus gave any signal. All of the streptomycetes that gave positive signals were producers of clavam-type metabolites (Elson et al., 1987)

Example 3

Disruption of the genomic cla gene

A 2.0 kb NcoI fragment that contained the entire cla gene was digested at its unique KpnI site and the ends

bioassay procedures described previously (Jensen et al. (1982), supra).

All of the resulting colonies with disrupted cla genes grew equally well on minimal medium and complex media and produced as much penicillin and cephamycin as did the wild-type, but produced no clavulanic acid (data not shown). HPLC analysis of cell supernatants confirmed the inability of the disrupted cla mutants to synthesize any clavulanic acid (data not shown).

Example 4

Protoplast formation and transformation

E. coli competent cell preparation and transformation were as described by Sambrook et al., (1989). Protoplasts of S. clavuligerus were, prepared, transformed and regenerated as described by Bailey et al. (1984), Bio/Technology, v. 2, pp. 808-811, with the following modifications. Dextrin and arginine in the regeneration medium were replaced by starch and sodium glutamate respectively. Protoplasts were heat shocked at 43°C for 5 min prior to the addition of DNA. Standard procedures were used for protoplasting and transformation of S. lividans (Hopwood et al. (1985)).

The 11.6 kb EcoR1 fragment from K6L2 (Fig. 9) was cloned into the EcoR1 site of pCAT-119. pCAT-119 is derivative of pUC119 which was prepared by insertionally inactivating the ampicillin resistance gene of pUC119 by the insertion of a chloramphenicol acetyltransferase gene (Jensen et al. (1989), Genetics & Molec. Biol. of Ind. Microorg., pp. 239-245 Ed. Hershberger, Amer. Soc. Microbiol). The PCAT-119 plasmid carrying the 11.6 kb fragment was then digested with PstI and ligated to the Streptomyces plasmid pIJ702, which had also been digested with PstI. The resulting bifunctional plasmid carrying the 11.6kb insert was capable of replicating in either E. coli (with selection for chloramphenicol resistance) or in S. lividans (with selection for thiostrepton

Example 5

Sequencing of 15 kb DNA fragment

Ordered sets of deletions were generated as described in Example 1 using fragments of the DNA insert from the cosmid clone K6L2 (Figure 9) and subcloned into the E. coli plasmids pUC118 andpUC119. Overlapping fragments were chosen which extended from the end of the pcbC gene downstream for a distance of about 15 kb ending at the BglII site. The deletion generated fragments were sequenced in both orientations as described in Example 1. The sequence is shown in Figure 2.

The present invention is not limited to the features of the embodiments described herein, but includes all variations and modifications within the scope of the claims.

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. An isolated genomic DNA molecule comprising the nucleotide sequence of Figure 2 (Sequence ID No.:1).

2. An isolated DNA molecule having the nucleotide sequence of nucleotides 2033 to 13636 of Figure 2 (Sequence ID No.:20).

3. An isolated DNA molecule having the nucleotide sequence of nucleotides 109 to 1764 of Figure 2 (Sequence ID No.:21).

4. An isolated DNA molecule having the nucleotide sequence of nucleotides 2216 to 3937 of Figure 2 (Sequence ID No.:22).

5. An isolated DNA molecule having the nucleotide sequence of nucleotides 3940 to 5481 of Figure 2 (Sequence ID No.:23).

6. An isolated DNA molecule having the nucleotide sequence of nucleotides 5654 to 6595 of Figure 2 (Sequence ID No.:24).

7. An isolated DNA molecule having the nucleotide sequence of nucleotides 6611 to 7588 of Figure 2 (Sequence ID No.:25).

8. An isolated DNA molecule having the nucleotide sequence of nucleotides 7895 to 9076 of Figure 2 (Sequence ID No.:26).

18. An isolated DNA molecule comprising a nucleotide sequence encoding the amino acid sequence of Figure 15.

19. An isolated DNA molecule comprising a nucleotide sequence encoding the amino acid sequence of Figure 16.

20. An isolated DNA molecule comprising a nucleotide sequence encoding the amino acid sequence of Figure 17.

21. An isolated DNA molecule comprising a nucleotide sequence encoding the amino acid sequence of Figure 18.

22. An isolated DNA molecule comprising a nucleotide sequence encoding the amino acid sequence of Figure 19.

23. An isolated protein having the amino acid sequence of Figure 10.

24. An isolated protein having the amino acid sequence of Figure 11.

25. An isolated protein having the amino acid sequence of Figure 12.

26. An isolated protein having the amino acid sequence of Figure 13.

27. An isolated protein having the amino acid sequence of Figure 14.

28. An isolated protein having the amino acid sequence of Figure 15.

transforming the host with a DNA molecule comprising a nucleotide sequence encoding one or more of the enzymes of the clavulanate synthetic pathway.

FIGURE 2 - 1

```
        |   10      |   20      |   30      |   40      |   50      |   60
  1 gcggaaccgg ccgcccctga gcggggcggc cgggaaggaa acgggccggt cgtcccctcg 60
                                                    <--End of ORF 1
 61 ggaggggcg gccggcccgt ccggtgcgcg cggtgggtgc ggcgcgggTC AGCCGGCCGC 120

121 GAGGTTGCTG AGGAACTTCG CGGCGACGGG GCCCGCGTCG GCGCCGCCCG ACCCGCCGTC 180

181 CTCCAGCAGG ACCGACCAGG CGATGTTCCG GTCGCCCTGG TAGCCGATCA TCCAGGCGTG 240

241 CGTCTTCGGC GGCTTCTCGG TGCCGAACTC GGCGGTACCG GTCTTGGCGT GCGGCTGTCC 300

301 GCCGAGGCCC CGCAGGGCGT CGCCGGCGCC GTCGGTGACG GTCGAACGCA TCATGGAACG 360

361 CAGCGAGTCG ACGATGCCCG GGGCCATCCG GGGGGCCTGG TGCGGCTTCT TGACCGCGTC 420

421 GGGCACCAGC ACGGGCTGCT TGAACTCGCC CTGCTTGACG GTGGCGGCGA TGGAGGCCAT 480

481 CACCAGGGGC GACGCCTCGA CCCTGGCCTG TCCGATGGTG GACGCGGCCT TGTCGTTCTC 540

541 GCTGTTGGAG ACGGGGACGC TGCCGTCGAA GGTGGAGGCG CCGACGTCCC AGGTGCCGCC 600

601 GATGCCGAAG GCTTCGGCGG CCTGCTTCAG GCTGGACTCG GAGAGCTTGC TGCGGGAGTT 660

661 GACGAAGAAC GTGTTGCAGG AGTGGGCGAA GCTGTCCCGG AAGGTCGAGC CCGCGGGCAG 720

721 CGTGAACTGG TCCTGGTTCT CGAAGCTCTG GCCGTTGACA TGGGCGAACT TCGGGCAGTC 780

781 GGCCCGCTCC TCCGGGTTCA TCCCCTGCTG GAGCAGGGCC GCGGTGGTGA CCACCTTGAA 840

841 GGTGGAGCCG GGCGGGTAGC GGCCCTCCAG CGCGCGGTTC ATGCCGGAGG GCACGTTCGC 900

901 GGCGGCCAGG ATGTTGCCGG TGGCGGGGTC GACGGCGACG ATCGCCGCGT TCTTCTTCGA 960

961 GCCCTCCAGG GCCGCCGCGG CGGCGGACTG GACCCGCGGG TCGATGGTGG TCTTCACCGG 1020

1021 CTTGCCCTCG GTGTCCTTGA GGCCGGTGAG CTTCTTGACC ACCTGGCCGG ACTCACGGTC 1080

1081 CAGGATCACG ACCGAGCGCG CCGCGCCGGA GCCGCCGGTG AGCTGCTTGT CGTAGCGGGA 1140

1141 CTGGAGGCCC GCCGAGCCCT TGCCGGTCCT GGGGTCGACC GCGCCGATGA TGGAGGCGGC 1200

1201 CTGGAGGACA TTGCCGTTGG CGTCGAGGAT GTCCGCGCGC TCCCGCGACT TGAGGGCGAG 1260

1261 GGTCTGCCCC GGAACCATCT GCGGATGGAT CATCTCGGTG TTGAACGCGA CCTTCCACTC 1320

1321 CTTGCCGCCG CCGACGACCT TCGCGGTGGA GTCCCAGGCG TACTCCCCGG CCCCGGGGAG 1380

1381 GGTCATTCTG ACGGTGAACG GTATCTCCAC CTCGCCCTCG GGGTTCTTCT CCCCGGTCTT 1440

1441 GGCGGTGATC TCCGTCTTCG TCGGCTTGAG GTTGGTCATG ACGGATTTGA TCAGCGACTC 1500

1501 GGCGTTGTCC GGGGTGTCCG TCAGCCCGGC GGCCGTCGGG GCGTCGCCCT TCTCCCAGGC 1560
```

FIGURE 2 - 3

```
3241 CGTGGAGCAC TTCGAGACCG CGACCGCCTC  CTTCGGGGCC AAGCAGCGCC ACGACATCGA 3300
3301 GCCGCTGCGC GCCCGGATCG CGGAGTTCCT  GGCCGACCCG GAGACCTACG AGGACGGCAT 3360
3361 GCGCGTCCAC CAGGTCATCG ACTCCATGAA  CACCGTCATG GAGGAGGCCG CCGAGCCCGG 3420
3421 CGAGGGCACG ATCGTCTCCG ACATCGGCTT  CTTCCGTCAC TACGGTGTGC TCTTCGCCCG 3480
3481 CGCCGACCAG CCCTTCGGCT TCCTCACCTC  GGCGGGCTGC TCCAGCTTCG GCTACGGCAT 3540
3541 CCCCGCCGCC ATCGGCGCCC AGATGGCCCG  CCCGGACCAG CCGACCTTCC TCATCGCGGG 3600
3601 TGACGGCGGC TTCCACTCCA ACAGCTCCGA  CCTGGAGACC ATCGCCCGGC TCAACCTGCC 3660
3661 GATCGTGACC GTCGTCGTCA ACAACGACAC  CAACGGCCTG ATCGAGCTGT ACCAGAACAT 3720
3721 CGGTCACCAC CGCAGCCACG ACCCGGCGGT  CAAGTTCGGC GGCGTCGACT TCGTCGCGCT 3780
3781 CGCCGAGGCC AACGGTGTCG ACGCCACCCG  CGCCACCAAC CGCGAGGAGC TGCTCGCGGC 3840
3841 CCTGCGCAAG GGTGCCGAGC TGGGTCGTCC  GTTCCTCATC GAGGTCCCGG TCAACTACGA 3900
                                  End of ORF 2--> Beginning of ORF 3-->
3901 CTTCCAGCCG GGCGGCTTCG GCGCCCTGAG  CATCTGAtcA TGGGGGCACC GGTTCTTCCG 3960
3961 GCTGCCTTCG GGTTCCTGGC CTCCGCCCGA  ACGGGCGGGG GCCGGGCCCC CGGCCCGGTC 4020
4021 TTCGCGACCC GGGGCAGCCA CACCGACATC  GACACGCCCC AGGGGGAGCG CTCGCTCGCG 4080
4081 GCGACCCTGG TGCACGCCCC CTCGGTCGCG  CCCGACCGCG CGGTGGCGCG CTCCCTCACC 4140
4141 GGCGCGCCCA CCACCGCGGT GCTCGCCGGT  GAGATCTACA ACCGGGACGA ACTCCTCTCC 4200
4201 GTGCTGCCCG CCGGACCCGC GCCGGAGGGG  GACGCGGAGC TGGTCCTGCG GCTGCTGGAA 4260
4261 CGCTATGACC TGCATGCCTT CCGGCTGGTG  AACGGGCGCT TCGCGACCGT GGTGCGGACC 4320
4321 GGGGACCGGG TCCTGCTCGC CACCGACCAC  GCCGGTTCGG TGCCGCTGTA CACCTGTGTG 4380
4381 GCGCCGGGCG AGGTCCGGGC GTCCACCGAG  GCCAAGGCGC TCGCCGCGCA CCGCGACCCG 4440
4441 AAGGGCTTCC CGCTCGCGGA CGCCCGCCGG  GTCGCCGGTC TGACCGGTGT CTACCAGGTG 4500
4501 CCCGCGGGCG CCGTGATGGA CATCGACCTC  GGCTCGGGCA CCGCCGTCAC CCACCGCACC 4560
4561 TGGACCCCGG GCCTCTCCCG CCGCATCCTG  CCGGAGGGCG AGGCCGTCGC GGCCGTGCGG 4620
4621 GCCGCGCTGG AGAAGGCCGT CGCCCAGCGG  GTCACCCCCG GCGACACCCC GTTGGTGGTG 4680
4681 CTCTCCGGCG GAATCGACTC CTCCGGGGTC  GCGGCCTGTG CGCACCGGGC GGCCGGGGAA 4740
4741 CTGGACACGG TGTCCATGGG CACCGACACG  TCCAACGAGT TCCGCGAGGC CCGGGCGGTC 4800
4801 GTCGACCATC TGCGCACCCG GCACCGGGAG  ATCACCATCC CGACCACCGA GCTGCTGGCG 4860
```

FIGURE 2 - 5

```
                                                        End of ORF 4-->
6541 GATCGGTGCG GAACTGCTCT ACCAGTACGC  CCGAGCCCAC AGAACCCAGT TGTGAaggag 6600
                Beginning of ORF 5-->
6601 acatcgtgtc ATGGCCTCTC CGATAGTTGA  CTGCACCCCG TACCGCGACG AGCTGCTCGC 6660

6661 GCTCGCCTCC GAGCTTCCCG AGGTGCCGCG  CGCGGACCTC CATGGCTTCC TCGACGAGGC 6720

6721 GAAGACGCTG GCCGCCCGTC TCCCGGAGGG  GCTGGCCGCC GCTCTCGACA CCTTCAACGC 6780

6781 CGTGGGCAGC GAGGACGGTT ATCTGCTGCT  GCGCGGGCTG CCCGTCGACG ACAGCGAGCT 6840

6841 GCCCGAGACG CCGACCTCCA CCCCGGCCCC  GCTGGACCGC AAGCGGCTGG TGATGGAGGC 6900

6901 CATGCTCGCG CTGGCCGGCC GCCGGCTCGG  TCTGCACACG GGGTACCAGG AGCTGCGCTC 6960

6961 GGGCACGGTC TACCACGACG TGTACCCGTC  GCCCGGCGCG CACTACCTGT CCTCGGAGAC 7020

7021 CTCCGAGACG CTGCTGGAGT TCCACACGGA  GATGGCGTAC CACATCCTCC AGCCGAACTA 7080

7081 CGTCATGCTG GCCTGCTCCC GCGCGGACCA  CGAGAACCGG GCGGAGACGC TGGTCGGCTC 7140

7141 GGTCCGCAAG GCGCTGCCCC TGCTGGACGA  GAAGACCCGG GCCCGTCTCT TCGACCGCAA 7200

7201 GGTGCCCTGC TGCGTGGACG TGGCCTTCCG  CGGCGGGGTC GACGACCCGG GCGCGATCGC 7260

7261 CAACGTCAAG CCGCTCTACG GGGACGCGAA  CGACCCGTTC CTCGGGTACG ACCGCGAGCT 7320

7321 GCTGGCGCCG GAGGACCCCG CGGACAAGGA  GGCCGTCGCC CATCTGTCCC AGGCGCTCGA 7380

7381 CGATGTGACC GTCGGGGTGA AGCTCGTCCC  CGGTGACGTC CTCATCATCG ACAACTTCCG 7440

7441 CACCACGCAC GCGCGGACGC CGTTCTCGCC  CCGCTGGGAC GGGAAGGACC GCTGGCTGCA 7500

7501 CCGCGTCTAC ATCCGCACCG ACCGCAATGG  ACAGCTCTCC GGCGGCGAGC GCGCGGGCGA 7560
                End of ORF 5-->
7561 CACCATCTCG TTCTCGCCGC GCCGCTGAgc  ccggctcccc gaggccctgg gccccggcgc 7620

7621 cggaaccggc tcccggtcct gccccctcac  ccgccgcgcg ggtgaggggg caggcccctt 7680

7681 tgtgccgggt gccgtgcgtc ctgcgagggt  gccggggcgg gggggacggc ggaggtgccc 7740

7741 ggcggccggg tgccgtgcgc cgcccgtggg  tgctgtacag cactccgtgt gccgtgcgcc 7800

7801 accccgtgca taaatttgcc actctatggg  aaataatgca gagtgcgacg ggtgaggccg 7860
                                            Beginning of ORF 6-->
7861 tcgccgtgcc ctttccgtga caggagacgc  tgacATGTCC GACAGCACAC CGAAGACGCC 7920

7921 CCGGGGATTC GTGGTGCACA CGGCGCCGGT  GGGCCTGGCC GACGACGGCC GCCACGACTT 7980

7981 CACCGTCCTC GCCTCCACCG CCCCGGCCAC  CGTGAGCGCC GTCTTCACCC GCTCCCGCTT 8040

8041 CGCCGGGCCG AGCGTCGTGC TGTGCCGGGA  GGCGGTGGCC GACGGGCAGG CGCGCGGTGT 8100

8101 GGTGGTGCTG GCCCGCAACG CGAATGTCGC  GACCGGCCTG GAGGGCGAGG AGAACGCGCG 8160
```

FIGURE 2 - 7

```
9841 TACCGGCTGC GGCCCGTGGC GACCGGCCCG  TACCGGATCG TCTCGTACAC CCGGGGCGAG 9900
9901 CTGGCCGTCC TGGAGCCCAA TCCGCACTGG  GACCCCGAGA CCGACCCGGT GCGCGTCCAG 9960
9961 CGCGCCTCCC GGATCGAGGT GCACCTCGGC  AAGGACCCGC ACGAGGTGGA CCGCATGCTG 10020
10021 CTGGCGGGCG AGGCCCATGT GGACCTCGCG  GGCTTCGGTG TGCAGCCCGC GGCCCAGGAG 10080
10081 CGCATCCTCG CCGAGCCGGA GCTGCGCGCG  CACGCGGACA ACCCGCTGAC CGGCTTCACC 10140
10141 TGGATCTACT GCCTGTCGAG CCGGATCGCC  CCGTTCGACA ATGTGCACTG CCGGCGGGCC 10200
10201 GTGCAGTTCG CCACCGACAA AGCGGCCATG  CAGGAGGCGT ACGGCGGCGC GGTGGGCGGC 10260
10261 GACATCGCGA CCACCCTGCT GCCCCCGACC  CTCGACGGCT ACAAGCACTT CGACCGCTAC 10320
10321 CCGGTCGGCC CCGAGGGCAC CGGCGACCTG  GAGGCCGCCC GCGCCGAGCT GAAGCTGGCC 10380
10381 GGGATGCCCG ACGGCTTCCG CACCAGGATC  GCCGCCCGCA AGGACCGGCT CAAGGAGTAC 10440
10441 CGGGCCGCCG AGGCGCTGGC CGCCGGGCTC  GCCCGGGTCG GCATCGAGGC GGAGGTGCTG 10500
10501 GACTTCCCGT CGGGCGACTA CTTCGACCGC  TACGGCGGCT GCCCGGAGTA TCTGCGCGAG 10560
10561 CACGGGATCG GGATCATCAT GTTCGGCTGG  GGCGCCGACT TCCCCGACGG ATACGGCTTC 10620
10621 CTCCAGCAGA TCACCGACGG GCGCGCGATC  AAGGAGCGCG GCAACCAGAA CATGGGCGAG 10680
10681 CTGGACGACC CGGAGATCAA CGCGCTGCTG  GACGAGGGGG CGCAGTGCGC CGACCCGGCG 10740
10741 CGGCGCGCGG AGATCTGGCA CCGCATCGAC  CAGCTCACGA TGGACCACGC GGTCATCGTT 10800
10801 CCGTATCTGT ACCCGCGGTC CTGCTCTAC  CGGCACCCGG ACACCCGCAA CGCCTTCGTC 10860
                                          End of ORF 7-->
10861 ACCGGCTCCT TCGGGATGTA CGACTACGTG  GCGCTCGGCG CGAAGTGAgc acggggtccg 10920
10921 gccccgggac cgtatgtccc ggggccggac  cccgccccgtt ccccgcccgg tccggtccgg 10980
                          <--End of ORF 8
10981 acccggtcgc ggccccgcTCA GCCGGACATC  CGGGCCCCGG CCGCGACCCC GCGCCGGATC 11040
11041 GGCCAGTGGC CCTGCGCCAG GGGCCGTTCC  ACGCTGCGGC AGGCGAGAGC GGCCTCGCGG 11100
11101 AACTCCGCCT CGTACAGCGC GAGCTGGCGC  AGGAACTGCC GGGTCGGGCC GGTCAGGCTG 11160
11161 GTCCCCCGCG GGCTGCGCAG CAGCAGCCGG  GCGCCGAGGG ACTGCTCCAG CCGGTGAATC 11220
11221 CGGCGGGTGA GCGCCGACTG GCTGATCGAC  AGCACCGCCG CGGCCCGGTT GATGCTGCCG 11280
11281 TGCCGGGCCA CGGCCTGGAG CAGATGGAGA  TCGTCCACAT CCAGTTTGCG GCCCTCGGCC 11340
11341 TGGCCGGGCA CGGAGCCCTG GTCGGGTCCC  GCCCCGAAGC GGCGGGCGTC CGCGCCGGTG 11400
11401 CGCTCCGCGT ACCACTGCGC CCACCAGGGC  TCGTCCAGCA GGTCGCGGTG GTGTTCGGCG 11460
11461 AAGCGCCGGA GCTGGACCTC GGCGATCAGC  GCGGCCAGCC GTCCCGCCAG CGCCCGGGGC 11520
```

FIGURE 2 - 9

```
13201 CCCGGCGGCG GTCAGCTCGT CACCCAGGGC  GCGCAGCTTC TCGACCCGGC GCGCGGCGAT 13260

13261 GGCCACGGCG GCGCCCTCGG CGGCCAGGGC  GCGGGCCGTG GCCTCGCCGA TGCCCGAGCT 13320
                                  <---beginning of ORF 9
13321 CGCGCCCGTG ATGAGCGCGA CTTTCCCCTG  GAGTGCGGAT GGCATcattt cctccacatg 13380

13381 gtgctgcgat cgtggtgagc gtatgaagaa  ggggtgagac ctgccgtgcc ggggcgggtt 13440

13441 ccgtacgccg gaccgttgcg gtgggcacgg  ccgaccgggt acggatggcc gcagttcccc 13500

13501 ggggagttcc cggggaatgg tgaataccgc  ggcgctctcc gatggtcttc ggaggacacc 13560

13561 cggggattca ccgggaatca gcggccggag  ttctccccgt ccacggcaga cgctatcagc 13620
                  *****  *
13621 gtcgcattcc ccggtgaatt cccttcggtg  gaccgggtta tgactgtttc cgccgggtta 13680

13681 tgcgcgccgc cccggcggac cggccacccg  cccgggggct gcggcagatt gggcgccacg 13740
                                  Beginning of ORF 10--->
13741 acatggcgcg agcagcgatc ggcggtggAT  GATGAACGAG GCAGCGCCTC AGTCCGACCA 13800

13801 GGTGGCACCG GCGTATCCGA TGCACCGGGT  CTGCCCGGTC GACCCGCCGC CGCAACTGGC 13860

13861 CGGGCTGCGG TCCCAGAAGG CCGCGAGCCG  GGTGACGCTG TGGGACGGCA GCCAGGTGTG 13920

13921 GCTGGTGACC TCGCACGCCG GGGCCCGGGC  CGTCCTGGGC GACCGCCGCT TCACCGCGGT 13980

13981 GACGAGCGCG CCCGGCTTCC CGATGCTGAC  CCGCACCTCC CAACTGGTGC GCGCCAACCC 14040

14041 GGAGTCGGCG TCGTTCATCC GCATGGACGA  CCCGCAGCAC TCCCGGCTGC GCTCGATGCT 14100

14101 CACCCGGGAC TTCCTGGCCC GCCGCGCCGA  GGCGCTGCGC CCCGCGGTGC GGGAGCTGCT 14160

14161 GGACGAGATC CTGGGCGGGC TGGTGAAGGG  GGAGCGGCCG GTCGACCTGG TCGCCGGACT 14220

14221 GACGATCCCG GTGCCCTCGC GGGTCATCAC  CCTGCTCTTC GGCGCCGGTG ACGACCGCCG 14280

14281 GGAGTTCATC GAGGACCGCA GCGCGGTCCT  CATCGACCGC GGCTACACCC CGGAGCAGGT 14340

14341 CGCCAAGGCC CGGGACGAAC TCGACGGCTA  TCTGCGGGAG CTGGTCGAGG AGCGGATCGA 14400

14401 GAACCCGGGC ACCGACCTGA TCAGCCGGCT  CGTCATCGAC CAGGTGCGGC CGGGGCATCT 14460

14461 GCGGGTCGAG GAGATGGTCC CGATGTGCCG  GCTGCTGCTG GTGGCCGGTC ACGGCACCAC 14520

14521 CACCAGCCAG GCGAGCCTGA GCCTGCTCAG  CCTGCTCACC GACCCGGAGC TGGCCGGGCG 14580

14581 CCTCACCGAG GACCCGGCCC TGCTGCCCAA  GGCGGTCGAG GAGCTGCTGC GCTTCCACTC 14640

14641 CATCGTGCAG AACGGGCTGG CCCGTGCCGC  GGTGGAGGAC GTCCAGCTCG ACGATGTGCT 14700

14701 CATCCGGGCG GGCGAGGGCG TGGTGCTGTC  GCTGTCGGCG GGCAACCGGG ACGAGACGGT 14760

14761 CTTCCCCGAC CCGGACCGGG TGGACGTGGA  CCGCGACGCC CGCCGCCATC TCGCCTTCGG 14820

14821 CCACGGCATG CACCAGTGCC TGGGCCAGTG  GCTGGCCCGG GTGGAGCTGG AGGAGATCCT 14880
```

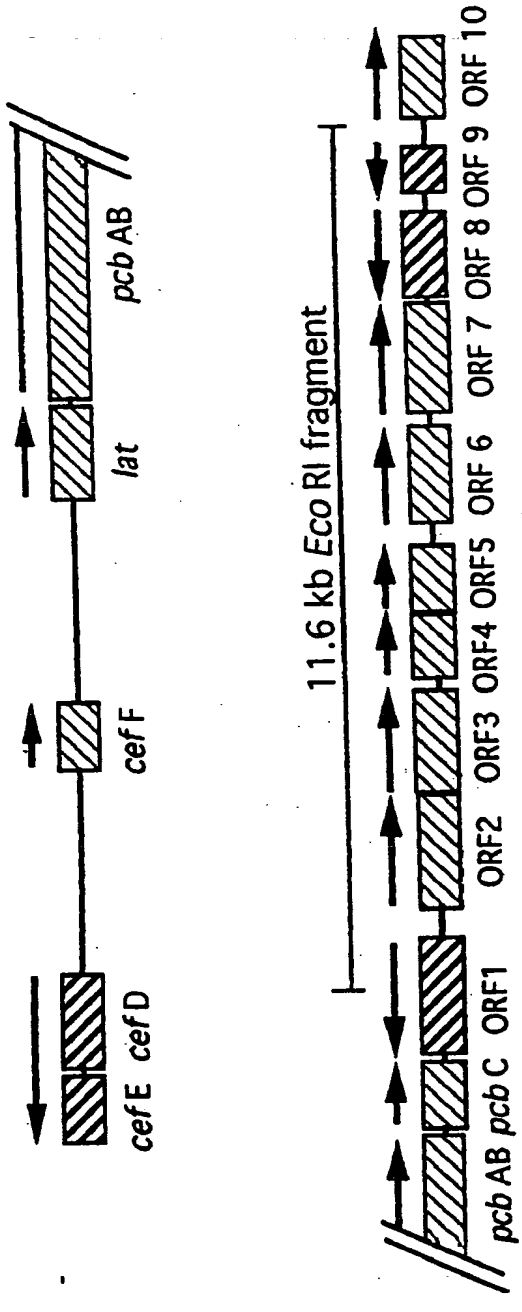

FIGURE 3

Simon M. Bourne

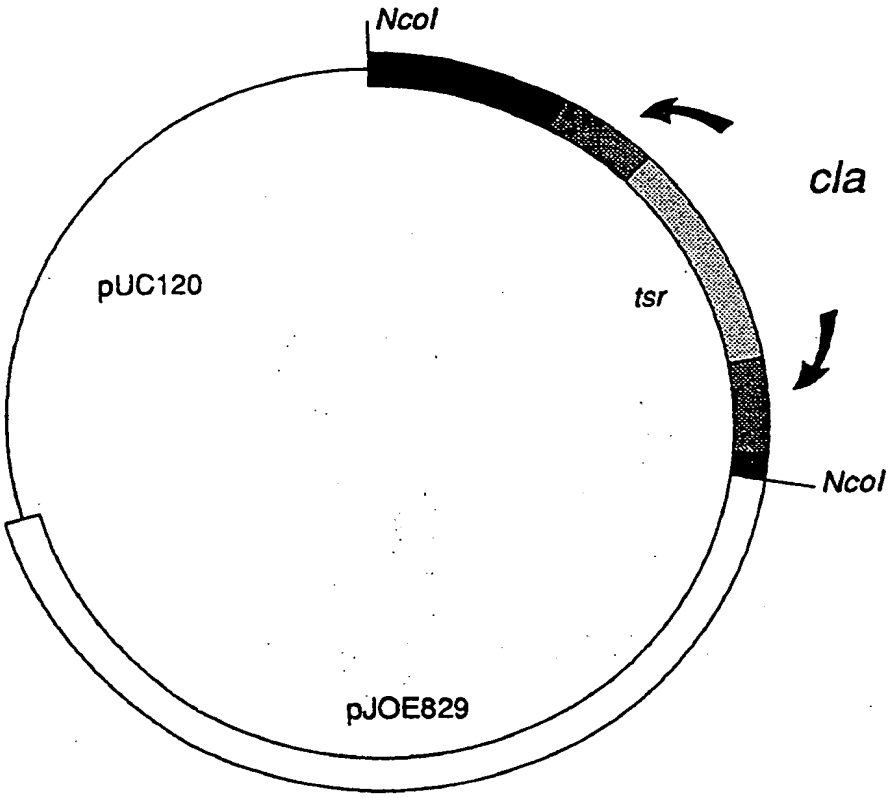

FIGURE 5

```
                                                                                  60
              1
S. Cl.  CLA    veridshvspryaqiptFmRLPhdpQPrgyDV--VviGaPyDggTSyRpGARfGPqAIR
E. co.  AUH   MSTIGhqYdNslvSnafGFIRLPmnfQPydsDadwVitGvPfDmaTSgRaGGRhGPaAIR
yeast ARG     MeT-GphY-NyyKnRelslvlAPFSgGQgkIGVEKGPkymIKhGL-qtsledlgwsteLE
rat ARG       MS---------sKpkplellGAPFSKGQPRGGVEKGPaaLRKAGL-------------vE
human ARG     MS---------aKSRtIGIIGAPFSKGQPRGGVEeGPTvLRKAGL-------------LE
                                                                                 120
              61
S. Cl.  CLA   seSgllhgvgidRgPgtFDI---INcVDaGDiNItpfDmniaidtaQsHISgLLKANaaf
E. co.  AUH   qvStnl-awehnRfPwnFDmrerINVVDcGDlvyafgDarEmSEkLQAHaeKLLaAGkrm
yeast ARG     psmdea-qfVgKlkmekdsttggssVmidGVKakRadIVGEAtklvynsVSKVvqANRfp
rat ARG       KLKEtE-ynV-rDhGDLaFvDvPNDSPFQIVKNPRS--VGKAnEQLAAvVAetqKNGtIS
human ARG     KLKEqE-cdV-KDyGDLpFaDiPNDSPFQIVKNPRS--VGKASEQLAgkVAqVkKNGRIS
                                                                                 180
              121
S. Cl.  CLA   LmiGGDHSLTvaaLRAVAeqhGpLAVVHIDAHsDTNpafyGgryhHGTpFrhgidEkLID
E. co.  AUH   LsfGGDHfvTIpILRAhAkhfGkmALVHfDAHTDTyan--GcefdHGTmFytapkEgLID
yeast ARG     LtLGGDHSiAIGtvSAVIdkyPDaGLIWIDAHaDINTi--esTpSGNLHGcPVSFLmgln
rat ARG       vVLGGDHSmAIGSISsHARVHPDLcVIWVDAHTDINTP--LTTsSGNLHGQPVaFLLKEL
human ARG     LVLGGDHSLAIGSISgHARVHPDLGVIWVDAHTDINTP--LTTtSGNLHGQPVSFLLKEL
                                                                                 240
              181
S. Cl.  CLA   PaamVQIGIRGHNPKPDSLdyarghGvrVvtAdefgelgVggtadLirekV---------
E. co.  AUH   PnhsVQIGIRt--------efdkdnGftVIdAcqvnDrsVddviaqvkqiV---------
yeast ARG     KdvphcpeslK--------WVpgnISpKklaYIGLRDVDaGEkkILKdLGIaaFSMyhVD
rat ARG       KGKfPDVPGFS--------WVTPCISAKDIVYIGLRDVDPGEHYIIKTLGIKYFSMTEVD
human ARG     KGKiPDVPGFS--------WVTPCISAKDIVYIGLRDVDPGEHYIIKTLGIKYFSMTEVD
                                                                                 300
              241
S. Cl.  CLA   --------------------GqRPVYvSvDIDvVDPAFAPGTGTPapGGLISREvLaLIR
E. co.  AUH   --------------------GdmPVYLtFDIDcLDPAFAPGTGTPViGGLTSdraIkLVR
yeast ARG     KyGInaViEmamkavhpetnGegPImcSyDVDGVDPlyIPATGTPVRGGLTIREGLfLVE
rat ARG       KLGIGKVME--ETfSYLLGRKKRPIHLSFDVDGLDPvFTPATGTPVVGGLsYREGLYITE
human ARG     rLGIGKVME--ETISYLLGRKKRPIHLSFDVDGLDPsFTPATGTPVVGGLTYREGLYITE
                                                                                 360
              301
S. Cl.  CLA   cv-gDLkpVGfDVMEVsPIYDhggITsI-------IATeigaELLYqyArahrTqIz
E. co.  AUH   gL-KDLNiVGmDVVEVaPaYDqseITal-------AAAtIALEmLYIqAaKkge
yeast ARG     rLaesGNLiaLDVVEcNPdLaihdIhVsnTisagcAIArcALGetII
rat ARG       EIYKTGLLSGLDIMEVNPTLGKTPEEVTRTVNTAVpITLsCFGtkREGNHKPeTDYLkPPK
human ARG     EIYKTGLLSGLDIMEVNPTLGKTPEEVTRTVNTAVAITLaCFGLaREGNHKP-IDYLnPPK
```

FIGURE 7

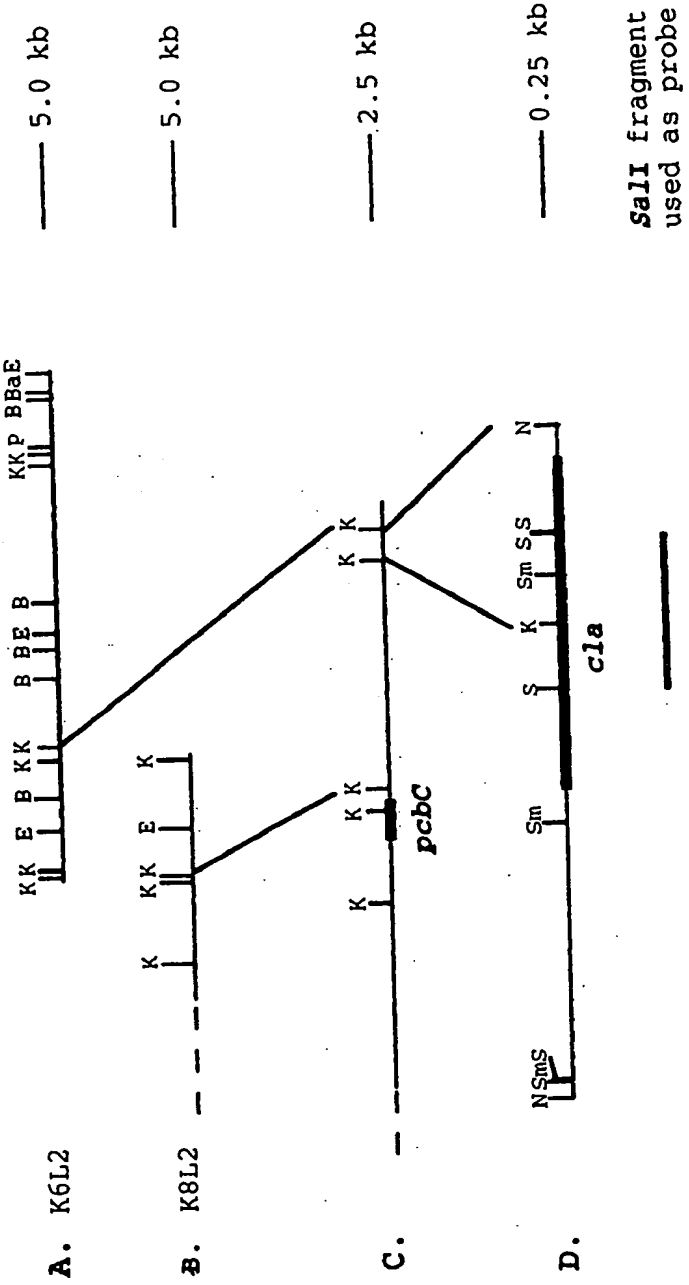

FIGURE 9

```
          |   10      |   20      |   30      |   40      |   50      |   60
  1 MSRVSTAPSG KPTAAHALLS RLRDHGVGKV FGVVGREAAS ILFDEVDPID FVLTRHEFTA 60
 61 GVAADVLARI TGRPQACWAT LGPGMTNLST GIATSVLDRS PVIALAAQSE SHDIFPNDTH 120
121 QCLDSVAIVA PMSLYAVELQ RPHEITDLVD SAVNAAMTEP VGPSFISLPV DLLGSSEGID 180
181 TTVPNPPANT PAKPVGVVAD GWQKAADQAA ALLAEAKHPV LVVGAAAIRS GAVPAIRALA 240
241 ERLNIPVITT YIAKGVLPVG HELNYGAVTG YMDGILNFPA LQTMFAPVDL VLTVGYDYAE 300
301 DLRPSMWQKG IEKKTVRISP TVNPIPRVYR PDVDVVTDVL AFVEHFETAT ASFGAKQRHD 360
361 IEPLRARIAE FLADPETYED GMRVHQVIDS MNTVMEEAAE PGEGTIVSDI GFFRHYGVLF 420
421 ARADQPFGFL TSAGCSSFGY GIPAAIGAQM ARPDQPTFLI AGDGGFHSNS SDLETIARLN 480
481 LPIVTVVVNN DTNGLIELYQ NIGHHRSHDP AVKFGGVDFV ALAEANGVDA TRATNREELL 540
541 AALRKGAELG RPFLIEVPVN YDFQPGGFGA LSIZ                           574
          |   10      |   20      |   30      |   40      |   50      |   60
```

FIGURE 11

```
        |   10      |   20      |   30      |   40      |   50      |   60
  1 VERIDSHVSP RYAQIPTFMR LPHDPQPRGY  DVVVIGAPYD GGTSYRPGAR FGPQAIRSES 60
 61 GLIHGVGIDR GPGTFDLINC VDAGDINLTP  FDMNIAIDTA QSHLSGLLKA NAAFLMIGGD 120
121 HSLTVAALRA VAEQHGPLAV VHLDAHSDTN  PAFYGGRYHH GTPFRHGIDE KLIDPAAMVQ 180
181 IGIRGHNPKP DSLDYARGHG VRVVTADEFG  ELGVGGTADL IREKVGQRPV YVSVDIDVVD 240
241 PAFAPGTGTP APGGLLSREV LALLRCVGDL  KPVGFDVMEV SPLYDHGGIT SILATEIGAE 300
301 LLYQYARAHR TQLZ                                                    314
        |   10      |   20      |   30      |   40      |   50      |   60
```

FIGURE 13

```
        |   10      |   20      |   30      |   40      |   50      |   60
  1 MSDSTPKTPR GFVVHTAPVG LADDGRHDFT VLASTAPATV SAVFTRSRFA GPSVVLCREA 60
 61 VADGQARGVV VLARNANVAT GLEGEENARE VREAVARALG LPEGEMLIAS TGVIGRQYPM 120
121 ESIREHLKTL EWPAGEGGFD RAARAIMTTD TRPKEVRVSV GGATLVGIAK GVGMLEPDMA 180
181 TLLTFFATDA RLDPAEQDRL FRRVMDRTFN AVSIDTDTST SDTAVLFANG LAGEVDAGEF 240
241 EEALHTAALA LVKDIASDGE GAAKLIEVQV TGARDDAQAK RVGKTVVNSP LVKTAVHGCD 300
301 PNWGRVAMAI GKCSDDTDID QERVTIRFGE VEVYPPKARG DQADDALRAA VAEHLRGDEV 360
361 VIGIDLAIAD GAFTVYGCDL TEGYVRLNSE YTTZ                             394
        |   10      |   20      |   30      |   40      |   50      |   60
```

FIGURE 15

```
      |   10      |   20      |   30      |   40      |   50      |   60
  1 MEVARRTGVR HGTVERRLDR LDRIVGLPLT  LRSRHTARLT TAGSRILVAG RRFFHQVDLA 60
 61 ARTHIFGHGS EAVDAPEVLS LVSTEPLLDE  VVEDAAASLD LLLSVRHEAP HQVAAQLAGY 120
121 QVDAAYTWSL QSPRHSLERS VRTCEVLDDP  LWVILPRDHP LAARREVSLA DLRDEIWVSE 180
181 TGPGSEILVT RVFQLAGLTA PTRLHITGAS  VARGILRRGD AIGLGSPTHP AVQDPSLVRR 240
241 SLAERPRRTT SLLVDPTIVP RALAGRLAAL  IAEVQLRRFA EHHRDLLDEP WWAQWYAERT 300
301 GADARRFGAG PDQGSVPGQA EGRKLDVDDL  HLLQAVARHG SINRAAAVLS ISQSALTRRI 360
361 HRLEQSLGAR LLLRSPRGTS LTGPTRQFLR  QLALYEAEFR EAALACRSVE RPLAQGHWPI 420
                                                                      433
421 RRGVAAGARM SGZ
      |   10      |   20      |   30      |   40      |   50      |   60
```

FIGURE 17

```
      |   10      |   20      |   30      |   40      |   50      |   60
  1 MMNEAAPQSD QVAPAYPMHR VCPVDPPPQL AGLRSQKAAS RVTLWDGSQV WLVTSHAGAR 60
 61 AVLGDRRFTA VTSAPGFPML TRTSQLVRAN PESASFIRMD DPQHSRLRSM LTRDFLARRA 120
121 EALRPAVREL LDEILGGLVK GERPVDLVAG LTIPVPSRVI TLLFGAGDDR REFIEDRSAV 180
181 LIDRGYTPEQ VAKARDELDG YLRELVEERI ENPGTDLISR LVIDQVRPGH LRVEEMVPMC 240
241 RLLLVAGHGT TTSQASLSLL SLLTDPELAG RLTEDPALLP KAVEELLRFH SIVQNGLARA 300
301 AVEDVQLDDV LIRAGEGVVL SLSAGNRDET VFPDPDRVDV DRDARRHLAF GHGMHQCLGQ 360
361 WLARVELEEI LAAVLRWMPG ARLAVPFEEL DFRHEVSSYG LGALPVTWZ           409
      |   10      |   20      |   30      |   40      |   50      |   60
```

FIGURE 19